**In the Claims**

1. (Currently amended) A method of gestural behavior analysis, comprising the steps of:

performing a training process using a combined audio/visual signal as a training data set, whereby prosodic audio features of said training data set are correlated with visual features of said training data set;

producing a statistical model based on results of said training process; and

applying said model to an actual data set to classify properties of gestural acts contained therein

wherein said training process comprises at least the steps of:

dividing said combined audio/visual signal into an audio component and a visual component;

identifying observable visual features of said visual component;

identifying observable prosodic features of said audio component; and

co-analyzing said audio and visual components to establish a correlation between said observable visual features and said observable prosodic features.

2. (Canceled)

3. (Currently amended) The method of claim [[2]] 1, wherein said training process further comprises at least the step of storing a database of reference gesture models, kinematical phases of gestural models, intonational representations of speech models, and combined gesture/speech models.

4. (Original) The method of claim 3, wherein said step of applying said model to an actual data set includes at least the steps of:

receiving an actual data set comprising a sequence of images and audio data corresponding to said sequence of images;

dividing said actual data set into an audio component and a visual component;

identifying observable visual features of said visual component of said actual data set;

identifying observable prosodic features of said audio component of said actual data set; and

comparing said identified observable visual and prosodic features of said visual and audio components of said actual data set, respectively, with said models stored in said database.

5. (Currently amended) The method of claim [[2]] 1, wherein said co-analyzing step comprises using a probabilistic framework to fuse gesture/speech co-occurrence information and visual gesture information to determine gesture occurrence in said actual data set.

6. (Original) The method of claim 5, wherein said probabilistic framework comprises a Bayesian framework.

7. (Currently amended) A system of gestural behavior analysis, comprising:

means for performing a training process using a combined audio/visual signal as a training data set, whereby prosodic audio features of said training data set are correlated with visual features of said training data set;

means for producing a statistical model based on results of said training process; and

means for applying said model to an actual data set to classify properties of gestural acts contained therein

wherein said training process comprises at least:

means for dividing said combined audio/visual signal into an audio component and a visual component;

means for identifying observable visual features of said visual component;

means for identifying observable prosodic features of said audio component; and

means for co-analyzing said audio and visual components to establish a correlation between said observable visual features and said observable prosodic features.

8. (Canceled)

9. (Currently amended) The system of claim [[8]] 7, wherein said training process further comprises at least means for storing a database of reference gesture models, kinematical phases of gestural models, intonational representations of speech models, and combined gesture/speech models.

10. (Original) The system of claim 9, wherein said means for applying said model to an actual data set includes at least:

means for receiving an actual data set comprising a sequence of images and audio data corresponding to said sequence of images;

means for dividing said actual data set into an audio component and a visual component;

means for identifying observable visual features of said visual component of said actual data set;

means for identifying observable prosodic features of said audio component of said actual data set; and

means for comparing said identified observable visual and prosodic features of said visual and audio components of said actual data set, respectively, with said models stored in said database.

11. (Currently amended) The system of claim [[8]] 7, wherein said means for co-analyzing comprises means for using a probabilistic framework to fuse gesture/speech co-occurrence information and visual gesture information to determine gesture occurrence in said actual data set.

12. (Original) The system of claim 11, wherein said probabilistic framework comprises a Bayesian framework.

13. (Currently amended) A computer program product for performing gestural behavior analysis, the computer program product comprising a computer-readable storage medium having computer-readable program code embodied in the medium, the computer-readable program code comprising:

computer-readable program code that performs a training process using a combined audio/visual signal as a training data set, whereby prosodic audio features of said training data set are correlated with visual features of said training data set;

computer-readable program code that produces a statistical model based on results of said training process; and

computer-readable program code that applies said model to an actual data set to classify properties of gestural acts contained therein

wherein said computer-readable program code that performs a training process comprises at least:

computer-readable program code that divides said combined audio/visual signal into an audio component and a visual component;

computer-readable program code that identifies observable visual features of said visual component;

computer-readable program code that identifies observable prosodic features of said audio component; and

computer-readable program code that co-analyzes said audio and visual components to establish a correlation between said observable visual features and said observable prosodic features.

14. (Canceled)

15. (Currently amended) The computer program product of claim [14] 13, wherein said computer-readable program code that performs a training process further comprises at least

computer-readable program code that stores a database of reference gesture models, kinematical phases of gestural models, intonational representations of speech models, and combined gesture/speech models.

16. (Original) The computer program product of claim 15, wherein said computer-readable program code that applies said model to an actual data set includes at least:

computer-readable code that receives an actual data set comprising a sequence of images and audio data corresponding to said sequence of images;

computer-readable code that divides said actual data set into an audio component and a visual component;

computer-readable code that identifies observable visual features of said visual component of said actual data set;

computer-readable code that identifies observable prosodic features of said audio component of said actual data set; and

computer-readable code that compares said identified observable visual and prosodic features of said visual and audio components of said actual data set, respectively, with said models stored in said database.

17. (Currently amended) The computer program product of claim [14] 13, wherein said computer-readable program code that co-analyzes comprises computer-readable program code that uses a probabilistic framework to fuse gesture/speech co-occurrence information and visual gesture information to determine gesture occurrence in said actual data set.

18. (Original) The computer program product of claim 17, wherein said probabilistic framework comprises a Bayesian framework.

19. (Original) A system for real-time continuous gesture recognition, comprising:

a) means for storing a database of reference gesture models, kinematical phases of gesture models, intonational representations of speech models, and combined gesture-speech models;

b) input means for receiving a sequence of images in real time, said images containing a gesticulating subject;

c) input means for receiving an audio signal from the gesticulating subject;

d) means for extracting a sequence of positional data of extremities of the subject;

e) means for extracting a pitch sequence of voice data points from the audio signal;

f) processing means for co-analyzing visual and acoustic signals from a recording; and

g) processing means for utilizing a probabilistic framework to fuse gesture-speech co-occurrence information and visual gesture information to determine gesture occurrence.

20. (Original) The system of claim 19 wherein said extremities comprise a head and hands of a subject.

21. (Original) The system of claim 19 wherein said probabilistic framework comprises a Bayesian framework.

22. (Original) A method for real-time continuous gesture recognition, comprising:

a) employing a database of reference gesture models, kinematical phases of gesture models, intonational representations of speech models, and combined gesture-speech models;

b) receiving a sequence of images in real time, said images containing a gesticulating subject;

c) receiving an audio signal from the gesticulating subject;

d) extracting a sequence of positional data of extremities of the subject;

e) extracting a pitch sequence of voice data points from the audio signal;

f) co-analyzing visual and acoustic signals from a recording by transforming the sequence of positional data extracted from each image to a sequence of velocity and acceleration features,

delimiting the sequence of the velocity and acceleration features through comparison to the kinematical phases of gestures,

extracting a set of acoustically prominent segments from the pitch sequence,

extracting a set of feature points from the acoustically prominent segments of pitch sequence,

extracting a set of feature points from the delimited kinematical phases of gestures represented by velocity and acceleration features of the extremities movement,

extracting a set of alignment measures of the feature points of pitch sequence and feature points of the extremities movement,

comparing the alignment measures of the feature points of pitch sequence and feature points of the extremities movement,

comparing the alignment measures of the feature points to co-occurrence gesture-speech models,

comparing the velocity and acceleration features of the extremities movement to the reference gesture models; and

g) utilizing a probabilistic framework to fuse gesture-speech co-occurrence information and visual gesture information to determine gesture occurrence.